

# PROJECT REPORT: Natural Disasters vs. GDP

## Team members:

- Jonas Brandenburg (if20b243), if20b243@technikum-wien.at
- Violeta Garcia Espin (if20b091), if20b091@technikum-wien.at
- Rawan Mousa (if20b129), [if20b129@technikum-wien.at](mailto:if20b129@technikum-wien.at)

**Date:** 14/11/2022.

**Repository:** [https://github.com/FreshJonas/Nat\\_Disasters\\_vs\\_GDP.git](https://github.com/FreshJonas/Nat_Disasters_vs_GDP.git)

## CONTENTS

Introduction.....	1
Methods .....	2
Results .....	2
Conclusions.....	2
4 levels of data handling.....	3
Big Data Criteria.....	4

## INTRODUCTION

Countries such as those in Southeast Asia and Central America seem to be more affected by natural disasters than the rest of the countries. Coincidentally, it is our subjective perception that these countries do not enjoy as much economic prosperity as other countries in more stable climates. The economic situation after Hurricane Irma in Haiti in 2017 or the Malay peninsula in Southeast Asia after the 2004 Tsunami are some examples. This perception is also shared by several authors [1] [2].

For this reason, we have carried out a small statistical analysis for our project in the subject “Big Data” about how a natural disaster can affect the economy of a country. Data from natural disasters from 1900 until 2021 was obtained from <https://www.emdat.be>. A country’s economy was evaluated according to its Gross National Product, which we collected through an API from the World Bank (<https://datahelpdesk.worldbank.org/knowledgebase/topics/125589-developer-information>).

For this we have used a dataset from EMD-dat. We have valued the country's economy according to its gross national product and we have collected it through an API of The World Bank data. Our hypothesis is that a natural disaster can negatively impact the GDP of a country.

## METHODS

The following tools and libraries have been used

Data obtaining:	Natural Disasters Dataset	CSV Download from <a href="https://www.emdat.be/">https://www.emdat.be/</a>
	Country GDP's	API call on World Bank for each country <a href="https://datahelpdesk.worldbank.org/knowledgebase/articles/898581-api-basic-call-structures">https://datahelpdesk.worldbank.org/knowledgebase/articles/898581-api-basic-call-structures</a>
Storing Data	Mapping	CouchDb Python library mapping <a href="https://couchdb-python.readthedocs.io/en/latest/">https://couchdb-python.readthedocs.io/en/latest/</a>
	NoSQL Database	CouchDB instance using Docker image and volume <a href="https://hub.docker.com/_/couchdb/">https://hub.docker.com/_/couchdb/</a>
Using / Analyzing Data	List of libraries in Jupyter Notebook 1	pandas → to deal with dataframes calendar, datetime, math → to deal with dates in the dataframe seaborn, matplotlib → to plot graphics ElementTree → to parse xml from API Genericpath → for relative paths
	List of libraries in Jupyter Notebook 2	pandas, numpy → to deal with data couchdb → to communicate with NoSQL database seaborn, matplotlib → to plot graphics scipy → for statistical analysis
	List of libraries in Jupyter Notebook 3	pandas, numpy → to deal with data couchdb → to communicate with NoSQL database sklearn → for machine learning
Collaboration Tools	Version Control	Public Github Repository at <a href="https://github.com/FreshJonas/Nat_Disasters_vs_GDP">https://github.com/FreshJonas/Nat_Disasters_vs_GDP</a>

## RESULTS

A positive correlation of 0.75 was shown between the 2019 GDP and the number of natural disasters since 1961. In addition, there seems to be a general increase in GDP regardless the presence of natural disasters among the countries that have suffered the greatest number of affected by natural disasters.

In addition, we use machine learning to try to predict the behavior of GDP based on natural disasters. Linear regression was used for that purpose, obtaining a mean absolute error in the prediction of 18.87%.

## CONCLUSIONS

The results do not corroborate our hypothesis, which can be due to various reasons. First, just one parameter regarding a country's economy was analyzed: the GDP. Possibly other indicators such as GINI or the GDP per capita could be affected by the presence of a natural disaster.

In addition, the increase in the GDP could be caused by other reasons, such as the normal growth of the economy in most countries. There are also various factors that can influence GDP and probably taking natural disasters alone as a predictor is not enough. In other words, in most countries, the presence of a natural disaster does not necessarily influence the GDP. This theory is supported by the poorly prediction of our linear regression algorithm.

Finally, this could be an example, that correlation and causality are not always equivalent. According to our data, a natural disaster does not necessarily mean that it will produce a change (positive or negative) in GDP.

This contradicts most of the consulted studies. For that reason, we believe that our hypothesis "Natural catastrophes can affect the economy of a country" might still be correct. However, we could not prove that in this study and more research is necessary for that. We believe that the reason for that is the little time available for the project and consequently the superficial treatment of the subject. A deeper study would require investigating various economic indicators, as well as clearly establishing measurement criteria, such as economic scores and severity of a disaster, to be able to adequately compare the countries among them

## 4 LEVELS OF DATA HANDLING

### 1. DATA SOURCE

We used external data for this project, and we did not generate our own data. We try to collect the data from recognized and reliable sources:

The natural disasters dataset was provided by EM-DAT, the Emergency Events Database. EM-DAT was created with the support of the World Health Organization (WHO) and the Belgian Government. It contains information about natural disasters from 1900 to 2021. According to EM-DAT, the data is compiled from various sources, including UN agencies, non-governmental organizations, insurance companies, research institutes and press agencies.

The GDP was requested through the World Bank API. The World Bank is an international financial institution that provides loans and grants to the governments of low- and middle-income countries for the purpose of pursuing capital projects. It also provides a free and open access to global development data. According to the World Bank Organization, most of the data comes from the statistical systems of member countries.

### 2. DATA STORAGE

The data was stored in a NoSQL database called couchDB. This database is characterized because it can handle huge amounts of unstructured and semi-structured data and it is used in places where internet access is not always guaranteed. We are aware that our data does not need such a powerful database and probably a relational database would be more suitable. However, a NoSQL database was a requirement for this project and all team members had already contact with this database since a presentation about it was made in this same subject.

### 3. DATA ANALYSIS

For the analysis, the data was collected from couchDB. MapReduce functions were used for those queries that need large iterations, for example for the number of different types of disasters. For other queries we have written specific functions in python.

Choosing the right variables and statistical methods has been challenging. Nonetheless, the correlation between GDP and number of disaster was measured using *corrcoef()* from the numpy library.

Finally, for machine learning linear regression was used. The mean absolute error was measured to evaluate the quality of the prediction.

#### 4. DATA OUTPUT

The data output is embedded in almost all phases, especially in the analysis phase. Because it is a small project with limited time, we have not been able to use more complex tools like Grafana or Tableau.

In addition, as our audience is composed of computer science students and teachers, we have considered that it is sufficient to use plots for data interpretation. The graphs have been generated with the matplotlib and seaborn python libraries.

## BIG DATA CRITERIA

### 1. Volume

The yearly GDP of countries as well as all natural disasters that are occurring on the entire planet pose a relatively small amount of data. Nevertheless, our data storing systems could handle much larger amounts of data. This is due to the fact that the CouchDB engine allows for using several database instances at different servers and providing a combined access interface. Furthermore, the fetched data from the World Bank-API gets uploaded directly to CouchDB without needing to store it locally which also enables a larger volume of data.

### 2. Velocity

The database is not big therefore the speed was high. However, we noticed that couchDB is not very fast when larger volumes of data are needed or with certain queries that require a more processing.

### 3. Variety

Our project deals with two different data types that are acquired independently from each other. The data received from the CSV file is of course structured. The data obtained from the World Bank-API also comes in a structured form. Nevertheless, since the json based CouchDB engine is used in the project any file that can be translated to json can be stored and processed by our data pipeline.

### 4. Veracity

Using an API that is provided by some external source always poses several problems regarding the veracity of data. There is no guarantee that the data values or structure will not change during the individual API calls. To tackle issues regarding the API response structure we parsed the XML data to Python Classes before uploading them to the database. This way we can prevent different formats in the database.

## 5. Value

The value of the data was determined when deciding our project idea. By definition, the collected data was valuable to our project since we were trying to observe correlations between the occurrence of natural disasters and the affected countries GDP's.

## REFERENCES

- [1] M. D. C. N. T. P. A.T. Bui, „The impact of natural disasters on household income, expenditure, poverty and inequality: evidence from Vietnam,“ *Appl. Econ.*, Bd. 46, Nr. 15, pp. 1751-1766, 2014.
- [2] C. V. C. D. Cappellia F, „The trap of climate change-induced “natural” disasters and inequality,“ *Global Environmental Change*, Bd. 70, 2021.